

Language identification setup basics

This quick reference guide contains a basic workflow for identifying the languages used in your documents. For more detailed information, see the Analytics section of the documentation site.

Language identification setup

The setup for running language identification is comprised of three components:

1. Saved search
2. Structured analytics set
3. View

1. Saved Search Setup

Use the following conditions and fields to create the saved search used for language identification. You do not need to set a sort order on this search.

- **Search Name**—there is no recommendation for the saved search name. Follow your team's normal protocol for naming searches.
- **Conditions**—extracted text size is greater than 0KB.
- **Fields**—any fields are acceptable.

2. Structured Analytics Set

Here are the steps and choices for creating a structured analytics set.

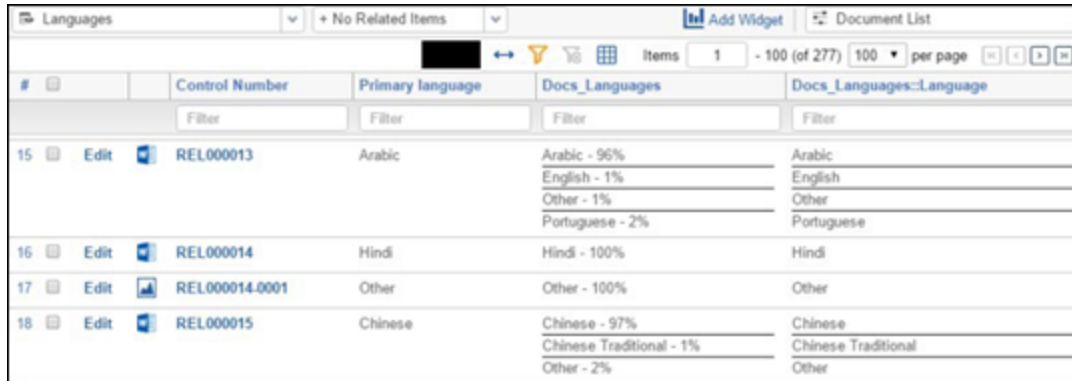
- **Name**—enter a name for the structured analytics set.
- **Prefix**—keep the default prefix or add your own prefix. Shorter prefixes (even just two characters, such as "LI") take up less space in your views.
- **Operations to run**—select Language identification.
- **Data source**—select the saved search you created above.

3. Language ID view

Once you run the structured analytics set, create the following view to see the results of your language identification operation.

- **Fields**
 - Control Number
 - [SAS Prefix]::Primary Language

- [SAS Prefix]::Docs_Languages
- [SAS Prefix]::Docs_Languages:Language
- **Conditions**—[SAS Prefix]::Primary Language is set
- **Sort**—you do not need a specific sort order for this view. However, you can create separate searches and views for each document set as well as widgets and dashboards.



The screenshot shows a web interface for document management. At the top, there's a header with 'Languages' and '+ No Related Items'. Below this is a table with columns: '#', 'Control Number', 'Primary language', 'Docs_Languages', and 'Docs_Languages:Language'. The table contains four rows of data, each representing a document with its control number, primary language, and a breakdown of detected languages.

#	Control Number	Primary language	Docs_Languages	Docs_Languages:Language
15	REL000013	Arabic	Arabic - 96% English - 1% Other - 1% Portuguese - 2%	Arabic English Other Portuguese
16	REL000014	Hindi	Hindi - 100%	Hindi
17	REL000014-0001	Other	Other - 100%	Other
18	REL000015	Chinese	Chinese - 97% Chinese Traditional - 1% Other - 2%	Chinese Chinese Traditional Other

Proprietary Rights

This documentation (“**Documentation**”) and the software to which it relates (“**Software**”) belongs to Relativity ODA LLC and/or Relativity’s third party software vendors. Relativity grants written license agreements which contain restrictions. All parties accessing the Documentation or Software must: respect proprietary rights of Relativity and third parties; comply with your organization’s license agreement, including but not limited to license restrictions on use, copying, modifications, reverse engineering, and derivative products; and refrain from any misuse or misappropriation of this Documentation or Software in whole or in part. The Software and Documentation is protected by the **Copyright Act of 1976**, as amended, and the Software code is protected by the **Illinois Trade Secrets Act**. Violations can involve substantial civil liabilities, exemplary damages, and criminal penalties, including fines and possible imprisonment.

©2024. Relativity ODA LLC. All rights reserved. Relativity® is a registered trademark of Relativity ODA LLC.